3

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

J. Sambrook
UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER

E.F. Fritsch
GENETICS INSTITUTE

T. Maniatis
HARVARD UNIVERSITY

Cold Spring Harbor Laboratory Press
1989

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

Printed in the United States of America

9 8 7 6 5 4 3 2

Book and cover design by Emily Harste

Cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. (Thomas R. Broker, Louise T. Chow, and James I. Garrels)

Cataloging in Publications data

Sambrook, Joseph
Molecular cloning : a laboratory manual / E.F. Fritsch, T. Maniatis—2nd ed.
p. cm.
Bibliography: p.
Includes index.
ISBN 0-87969-309-6
1. Molecular cloning—Laboratory manuals. 2. Eukaryotic cells--Laboratory manuals. I. Fritsch, Edward F. II. Maniatis, Thomas III. Title.
QH442.2.M26 1989
574.87'3224—dc19 87-35464

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724. Phone: 1-800-843-4388. In New York (516)367-8423.

found within both exons and introns of many eukaryotic genes. Such "cryptic" splice sites can be efficiently utilized when the normal splice sites are inactivated by mutation (Treisman et al. 1983; Wieringa et al. 1983).

Both the distance between splice sites and the DNA sequences surrounding them may influence the pathway of splicing in pre-mRNAs that contain multiple introns (Reed and Maniatis 1986). Alterations to the exon sequences flanking 5′ or 3′ splice sites can dramatically affect the efficiency with which the adjacent splice site is utilized. These findings are relevant to the design of eukaryotic expression vectors: Substitution of exon sequences or juxtaposition of normally noninteracting splice sites in a hybrid transcription unit might lead to the appearance of inappropriately spliced transcripts that cannot be translated.

Early studies of the expression of β-globin cDNA clones in cultured mammalian cells suggested that splicing is required for the production of cytoplasmic β-globin mRNA (Hamer and Leder 1979a,b,c). Furthermore, the expression of a gene with a mutation at a natural splice site could be rescued by insertion of a heterologous intron into the transcription unit (Gruss et al. 1979; Gruss and Khoury 1980). It is now known that this requirement for splicing signals is not absolute: Many cDNAs have been efficiently expressed from vectors that lack splicing signals (see, e.g., Gething and Sambrook 1981; Treisman et al. 1981). However, because the presence of an intron has proven to be deleterious in only a few cases and because some genes appear to be expressed more efficiently when introns are present, we recommend the use of vectors that contain a splice donor and acceptor site within the mammalian transcription unit.

ELEMENTS FOR REPLICATION AND SELECTION

In addition to the elements already described, eukaryotic vectors may contain other specialized elements intended to increase the level of expression of cloned genes or to facilitate the identification of cells that carry the transfected DNA.

Viral replicons

A number of animal viruses contain DNA sequences that promote the extrachromosomal replication of the viral genome in permissive cell types. Plasmids bearing these viral replicons are replicated episomally as long as the appropriate *trans*-acting factors are provided by genes either carried on the plasmid or within the genome of the host cell. Different viral replicons work with different efficiencies. Plasmid vectors containing the replicons of papovaviruses such as SV40 or polyomavirus replicate to extremely high copy number in cells that express the appropriate viral T antigen. Because the transfected cells die after 3 or 4 days, when the number of plasmid molecules exceeds 10^4 copies/cell, these systems are used for the transient, but abundant, expression of the transfected genes (see pages 16.17–16.22). Plasmid vectors containing replicons from viruses such as bovine papillomavirus (see pages 16.23–16.26) and Epstein-Barr virus (see pages 16.26–27) are propagated episomally at lower copy numbers (usually < 100 copies/cell) and do not generally cause cell death. These vectors can be used to isolate stable

lines of cells that permanently express more modest levels of the transfected genes.

Genes encoding selectable markers

DNA, which enters only a small proportion of mammalian cells in a given culture, becomes stably maintained in an even smaller fraction. In a very few cases—for example, when the cells are transformed by an oncogene—stably transfected cells can be identified because they express an altered phenotype such as morphological transformation, loss of contact inhibition, or increased growth rate. However, in the great majority of cases, isolation of cell lines that express the transfected gene is achieved by introduction into the same cells of a second gene that encodes a selectable marker, i.e., an enzymatic activity that confers resistance to an antibiotic or other drug. Some of the markers described below are dominant and can be used with any type of mammalian cell; others must be used with particular cell lines that lack the relevant enzyme activity.

In early experiments, the genes encoding the protein of interest and the selectable marker were included on a single vector. However, Wigler et al. (1979) found that mammalian cells capable of taking up DNA do so efficiently, so that two unlinked plasmids can be cotransfected with high frequency ($>90\%$). Cotransfection, which obviates the need to construct complex recombinants, has become the standard method of introducing a selectable marker (on one plasmid) and the gene of interest (on another plasmid) into mammalian cells. The selectable markers that are currently used include:

- *Thymidine kinase.* The thymidine kinase gene (*tk*), which is expressed in most mammalian cells, codes for an enzyme that is involved in the salvage pathway for synthesis of thymidine nucleotides. A number of tk^- cell lines have been isolated from different mammalian species, including mouse (Ltk^- cells) (Kit et al. 1963; Wigler 1977), human ($143tk^-$ cells) (Bacchetti and Graham 1977), and rat (Rat-2 fibroblast cells) (Topp 1981). These mutant cell lines, in contrast to their wild-type parents, will grow in medium that contains the thymidine analog 5-bromodeoxyuridine. Szybalska and Szybalski (1962) and Littlefield (1964, 1966) developed a selective medium containing hypoxanthine, aminopterin, and thymidine (HAT medium; see Appendix A) in which only cells expressing the *tk* gene will grow. By the appropriate use of this medium, it is therefore possible to select for or against cells that express the *tk* gene.

  Early cotransfection experiments utilized purified fragments of herpes simplex virus (HSV) DNA that contained the viral *tk* gene (Wigler et al. 1977). Subsequent cloning of the *tk* gene both from HSV (Colbère-Garapin et al. 1979) and from chicken cells (Perucho et al. 1980) made it possible to construct plasmids such as that shown in Figure 16.1A for use in cotransfection experiments. The primary limitation of these vectors is that they can be used only in tk^- cell lines.

- *Dihydrofolate reductase.* Mutants of CHO cells that lack the enzyme dihydrofolate reductase (Urlaub and Chasin 1980) cannot synthesize tetrahydrofolate and therefore can grow only in media supplemented with

thymidine, glycine, and purines. Transfection of these cells with vectors that express a cloned copy of the dihydrofolate reductase gene (*dhfr*) gives rise to clones that can grow in the absence of these supplements (Subramani et al. 1981; Kaufman and Sharp 1982a,b; Kaufman et al. 1985; see Figures 16.1B and 16.3C).

DHFR can be inhibited by methotrexate, a folate analog. Progressive selection of cells that are resistant to increasing concentrations of methotrexate leads to amplification of the *dhfr* gene, with concomitant amplification of extensive regions of the DNA that flank the *dhfr* sequences (Schimke 1982). DNAs that are cotransfected with the *dhfr* gene tend to become integrated into the same region of the cellular chromosome and therefore can frequently be coamplified with *dhfr*. Alternatively, cells lacking DHFR activity can be transfected with a recombinant construct containing the gene of interest linked to the *dhfr* gene. The linked gene is then amplified by selecting with successively higher concentrations of methotrexate. The resulting cell lines express very high levels of the desired recombinant protein product (Kaufman and Sharp 1982a,b; Kaufman et al. 1985). This approach is described in more detail on page 16.28.

The coamplification method has also been adapted for use with cells that synthesize wild-type levels of DHFR. In one approach, the *dhfr* gene was placed under the control of a strong promoter, thereby conferring on transfected cells the ability to grow in concentrations of methotrexate that would be lethal to cells expressing normal, wild-type levels of the enzyme (Murray et al. 1983). Alternatively, cells transfected with a plasmid that carries a dominant selectable marker (e.g., resistance to geneticin [G418]), the *dhfr* gene, and the gene of interest are selected first for their ability to grow in G418 and then for their ability to grow in progressively higher concentrations of methotrexate (Kim and Wold 1985). Finally, an altered form of the *dhfr* gene encoding an enzyme that is more resistant to methotrexate has been utilized as a dominant selectable marker for cotransformation experiments in a broad range of cell types (Spandidos and Siminovitch 1977; O'Hare et al. 1981; Simonsen and Levinson 1983).

Note: G418 is now commercially available. Because cultured lines of mammalian cells differ widely in their sensitivity to this antibiotic, the concentration appropriate for the selection of stably transfected cells must be determined empirically.

- *Aminoglycoside phosphotransferase*. The mostly widely used dominant selection system utilizes the bacterial gene encoding aminoglycoside 3′ phosphotransferase (APH). Two distinct APH enzymes, encoded by the bacterial transposons Tn*5* and Tn*601*, confer resistance to aminoglycoside antibiotics such as kanamycin, neomycin, and geneticin, which inhibit protein synthesis in both prokaryotic and eukaryotic cells. Eukaryotic cells do not normally express an endogenous APH activity, but they are capable of expressing the enzymes encoded by the bacterial transposons. When fused to eukaryotic transcriptional regulatory elements, the genes encoding APH can be used as dominant markers to select cells that take up exogenous DNA (Jimenez and Davies 1980; Colbère-Garapin et al. 1981). The first APH (*neo*^{r}) vectors designed for mammalian cells expressed the Tn*5* *neo*^{r} gene under the control of the HSV *tk* promoter and polyadenylation sequences (Colbère-Garapin et al. 1981). Subsequently, vectors were

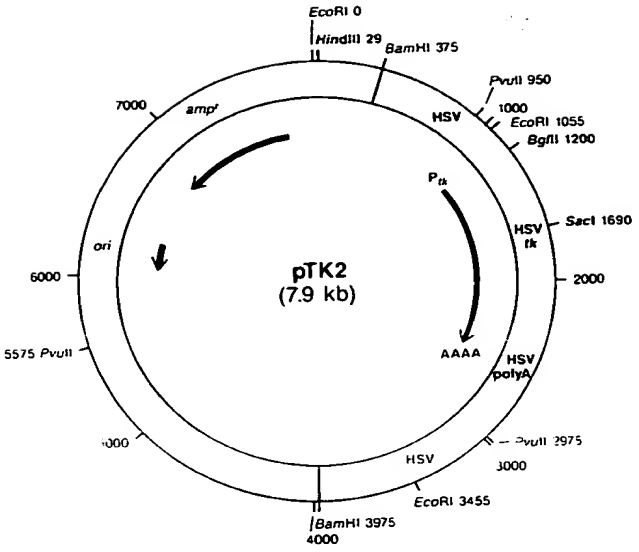

FIGURE 16.1A
pTK2 is a derivative of pBR322 that carries a 3.6-kb *Bam*HI fragment of herpes simplex virus (HSV) encoding thymidine kinase (*tk*). The positions of the *tk* promoter (P_{tk}) and the polyadenylation site (polyA; AAAA) are indicated.

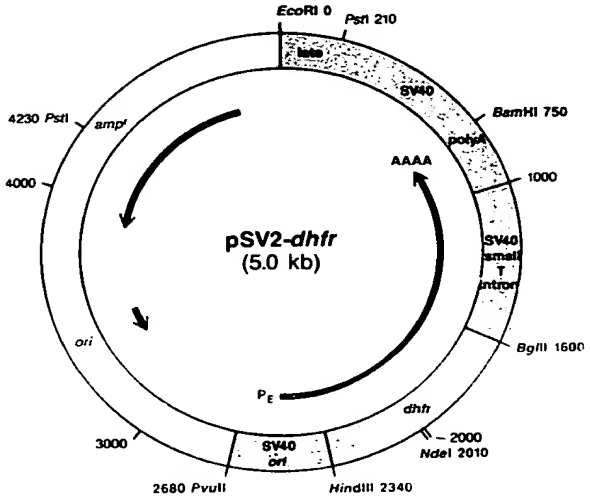

FIGURE 16.1B
pSV2-*dhfr* carries the SV40 origin (SV40 *ori*) and expresses dihydrofolate reductase (*dhfr*) from the SV40 early promoter (P_E). The SV40 small T intron and polyadenylation site (polyA; AAAA) are shown.

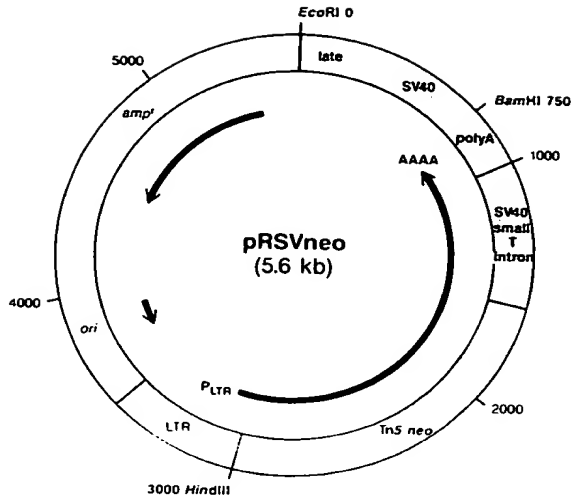

FIGURE 16.1C
pRSVneo expresses aminoglycoside phosphotransferase (APH) encoded by the bacterial transposon gene Tn*5* *neo*^{r} from the Rous sarcoma virus (RSV) LTR promoter (P_{LTR}). The SV40 small T intron and polyadenylation site (polyA; AAAA) are located downstream from Tn*5* *neo*.

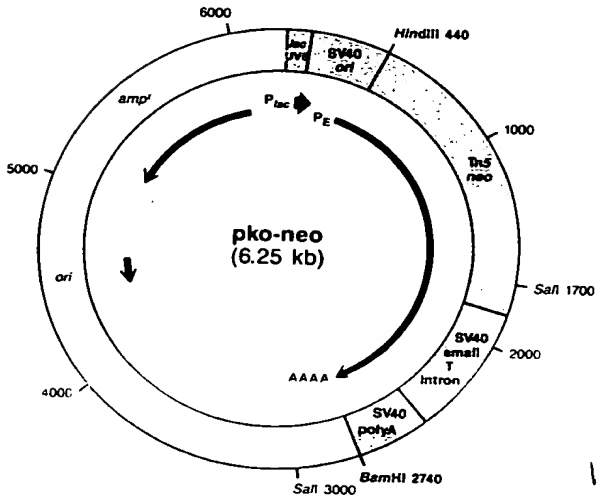

FIGURE 16.1D
pko-neo expresses aminoglycoside phosphotransferase encoded by the bacterial transposon gene Tn*5* *neo*^{r} from the eukaryotic SV40 early promoter (P_E) or the prokaryotic *E. coli lac*UV5 promoter (P_{lac}). The SV40 origin (SV40 *ori*), SV40 small T intron, and SV40 polyadenylation sites (polyA; AAAA) are present.

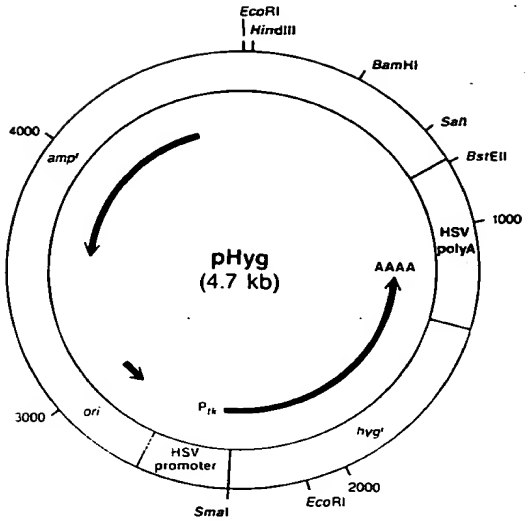

FIGURE 16.1E
pHyg directs the expression of the *E. coli* gene encoding hygromycin B phosphotransferase (hyg^r) using the herpes simplex virus promoter (P_{tk}) and polyadenylation site (HSV polyA; AAAA).

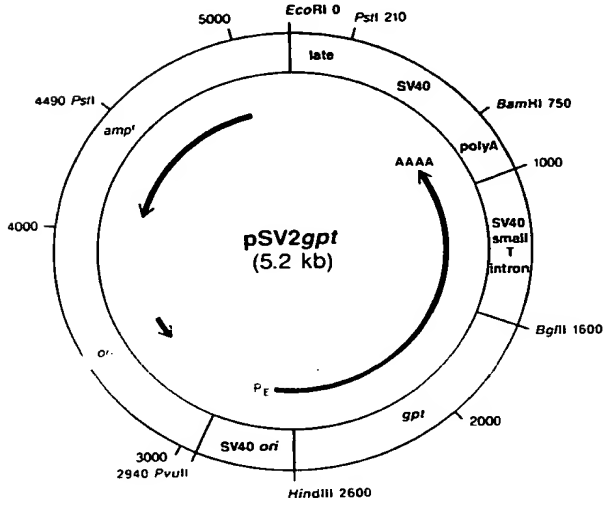

FIGURE 16.1F
In pSV2*gpt*, the *E. coli* xanthine-guanine phosphoribosyl transferase gene (*gpt*) is expressed using the SV40 early promoter (P_E) located in the SV40 origin (SV40 *ori*), the SV40 small T intron, and the SV40 polyadenylation site (polyA; AAAA).

developed that express the Tn*5* *neo*^{r} gene under the control of SV40 regulatory elements (Chia et al. 1982; Southern and Berg 1982; Okayama and Berg 1983; Van Doren et al. 1984). Vectors such as pSV2-*neo* (Southern and Berg 1982) and pRSVneo (Figure 16.1C), which have been widely used in cotransformation experiments, contain a version of the Tn*5* *neo*^{r} gene that retains prokaryotic promoter sequences between the eukaryotic promoter and the APH coding sequences. This configuration yields a vector that can confer antibiotic resistance upon both prokaryotic and eukaryotic cells. However, perhaps because the bacterial promoter contributes several upstream AUG codons, the efficiency of translation of APH mRNAs synthesized from these vectors is comparatively low in mammalian cells (Chen and Okayama 1987). Vectors such as pko-neo (Figure 16.1D) (Van Doren et al. 1984) and pcDneo (Okayama and Berg 1983; Chen and Okayama 1987), which lack prokaryotic promoter sequences, are therefore preferred.

- *Hygromycin B phosphotransferase.* The *E. coli* gene encoding hygromycin B phosphotransferase (Gritz and Davies 1983) can be used as a dominant selectable marker in much the same way as the APH gene. When the hygromycin B phosphotransferase gene (*hyg*) is introduced into mammalian cells on an appropriate expression vector (e.g., pHyg, Figure 16.1E) (Sugden et al. 1985), the transfected cells become resistant to the antibiotic hygromycin. Resistance to neomycin and to hygromycin can be selected for independently and simultaneously in cell lines that have been transfected with both genes. Thus, two different vectors can be introduced into one cell line, either simultaneously or sequentially.

- *Xanthine-guanine phosphoribosyl transferase.* The *gpt* gene of *E. coli* encodes the enzyme xanthine-guanine phosphoribosyl transferase (XGPRT), which is the bacterial analog of the mammalian enzyme hypoxanthine-guanine phosphoribosyl transferase (HGPRT). Whereas only hypoxanthine and guanine are substrates for HGPRT, XGPRT will also efficiently convert xanthine into XMP, which is a precursor of GMP. The bacterial *gpt* gene has been cloned and expressed in mammalian cells under the control of an SV40 promoter (Mulligan and Berg 1980, 1981a,b) (see, e.g., Figure 16.1F). Vectors expressing XGPRT restore the ability of mammalian cells lacking HGPRT activity to grow in HAT medium (Szybalska and Szybalski 1962; Littlefield 1964, 1966).

  Of much greater general use is the application of the *gpt* gene as a dominant selection system, which can be applied to any type of cell (Mulligan and Berg 1981a,b). Vectors expressing XGPRT confer upon wild-type mammalian cells the ability to grow in medium containing adenine, xanthine, and the inhibitor mycophenolic acid. Mycophenolic acid blocks the conversion of IMP into XMP and inhibits the de novo synthesis of GMP. The selection can be made more efficient by the addition of aminopterin, which blocks the endogenous pathway of purine biosynthesis.

- *CAD.* A single protein, CAD, possesses the first three enzymatic activities of de novo uridine biosynthesis (carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase). Transfection of vectors expressing the CAD protein from Syrian hamsters into CAD-deficient (UrdA) mutants of CHO cells allows selection of CAD$^+$ transfectants that are able to grow in the absence of uridine (Robert de Saint Vincent et al. 1981).

L-Phosphonacetyl-L-aspartate (PALA) is a specific inhibitor of the aspartate transcarbamylase activity of CAD. Growth of wild-type or transfected mammalian cells in the presence of increasing concentrations of PALA leads to the amplification of the CAD gene and DNA sequences linked to it (Kempe et al. 1976; Robert de Saint Vincent et al. 1981; Wahl et al. 1984). The *E. coli* gene encoding aspartate transcarbamylase (*pyrB*), when expressed in CHO cells deficient in aspartate transcarbamylase, is also amplified by PALA selection (Ruiz and Wahl 1986).

- *Adenosine deaminase*. Adenosine deaminase (ADA) is present in virtually all animal cells, but it is normally synthesized in minute quantities and is not essential for cell growth. However, because ADA catalyzes the irreversible conversion of cytotoxic adenine nucleosides to their respective nontoxic inosine analogs, cells propagated in the presence of toxic concentrations of adenosine or its analog 9-β-D-xylofuranosyl adenine (Xyl-A) require ADA for survival (for references and review, see Kaufman 1987). Under conditions where ADA is required for cell growth, amplification of the gene can be achieved in the presence of increasing concentrations of 2′-deoxycoformycin (dCF), a transition-state analog of adenine nucleotides that strongly inhibits the enzyme. In cells selected for their ability to resist high concentrations of 2′-deoxycoformycin, it has been shown that ADA was overproduced 11,400-fold and represented 75% of the soluble protein synthesized by the cells (Ingolia et al. 1985).
- *Asparagine synthetase*. The *E. coli* gene coding for asparagine synthetase (AS) is a potentially useful, dominant, amplifiable marker for mammalian cells. Because the bacterial enzyme uses ammonia as an amide donor—in contrast to the mammalian enzyme, which uses glutamine—cells that express the bacterial AS gene will grow in asparagine-free medium containing the glutamine analog albizziin. Subsequently, the transfected AS gene can be amplified by selection in medium containing increasing concentrations of β-aspartyl hydroxamate, an analog of aspartic acid.

Foreign DNA Sequences

DNAs encoding the foreign protein of interest are usually cloned as cDNAs that lack all of the controlling elements required for expression in mammalian cells but may contain ancillary sequences introduced during the construction of the cDNA library (e.g., homopolymeric stretches of guanine or cytosine residues, synthetic linkers, etc.). No consensus exists as to whether or not these ancillary sequences need to be removed before the cDNA can be expressed in mammalian cells. However, since such sequences never enhance, and in some circumstances may suppress, the level of expression of foreign DNAs in mammalian cells (Simonsen et al. 1982), most workers prefer to remove as many extraneous sequences as is conveniently possible. Less frequently, DNAs encoding the foreign protein of interest are obtained as a genomic copy in which the coding sequences may be interrupted by one or more introns. A complete genomic copy will have all the controlling sequences necessary for the expression of the protein in some, but not necessarily all, cell types. Because the specificity of these sequences determines the range of cell types in which the gene will be active, replacement